

## Supplemental

**Supplemental Table 1. The significant distinctive features by univariate analysis. The two features in bold were selected in the final model.**

Rank	Feature	Type	P-value	AUC	Correlation
1	<b>SD_IDM</b>	Texture	0.000038	0.83	-
2	<b>BB_AP</b>	Shape	0.00013	0.84	+
3	SD Short Run Emphasis	Texture	0.00021	0.81	-
4	BB_SI	Shape	0.00035	0.82	+
5	Weighted Principal Moments3	Shape+Intensity	0.00072	0.80	+
6	Equivalent Ellipsoid Diameter1	Shape	0.0019	0.79	+
7	SD Run Length Nonuniformity	Texture	0.0022	0.78	+
8	SD Grey Level Nonuniformity	Texture	0.0024	0.78	+
9	WeightedPrincipalMoments2	Shape+Intensity	0.0035	0.78	+
10	2D Roundness	Shape	0.0037	0.79	-
11	Mean Short Run Low Grey Level Emphasis	Texture	0.0064	0.77	-
12	Mean Short Run Emphasis	Texture	0.0067	0.77	-
13	SD Inertia	Texture	0.0079	0.77	-
14	SD Low Grey Level Run Emphasis	Texture	0.023	0.75	-

**Supplemental Table 2. Inter-observer variation in contouring the PN volume. The STAPLE-estimated sensitivity, specificity and Jaccard index measured the agreement between the individual contours and the group consensus contours.**

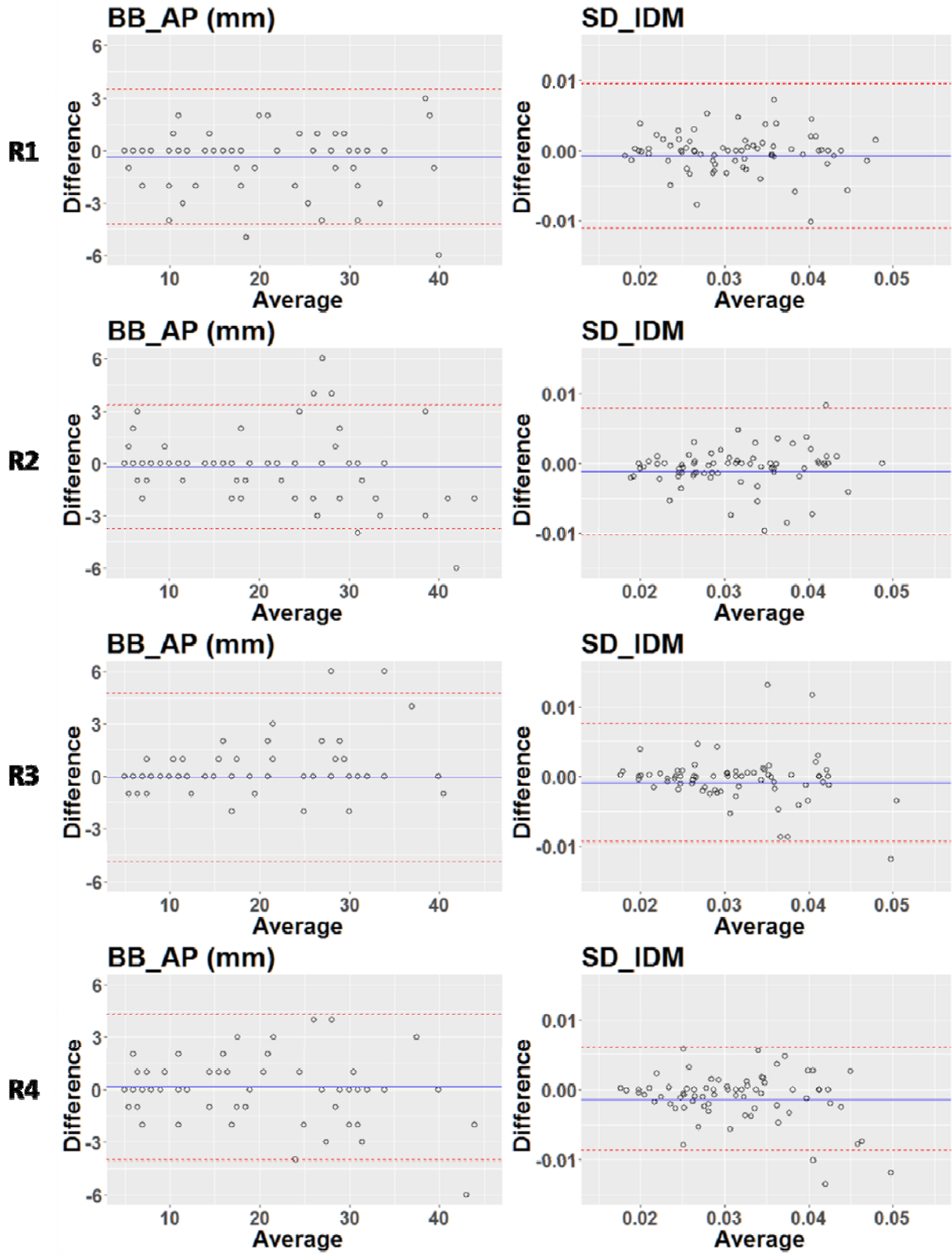
Radiologist	Sensitivity (%)	Specificity (%)	Jaccard (%)
R1	90.1 ± 10.8	85.9 ± 19.1	80.1 ± 12.4
R2	88.6 ± 11.8	90.0 ± 9.5	81.7 ± 11.2
R3	88.4 ± 10.8	93.8 ± 7.6	83.9 ± 10.4
R4	84.3 ± 14.3	91.3 ± 10.8	77.7 ± 13.5

**Supplemental Table 3. The intra-class correlation coefficient (ICC) among the individual contours and the group consensus contours.**

Feature	ICC	P-value
<b>BB_AP</b>	0.95	4.38E-100
<b>SD_IDM</b>	0.78	7.43E-47

**Supplemental Table 4. Comparison between the predictions by the SVM-LASSO model using features from the group consensus contours and each radiologist's contours; P-values were computed by McNamer's test between predictions based on the consensus and individual contours.**

<b>Radiologist</b>	<b>Sensitivity (%)</b>	<b>Specificity (%)</b>	<b>Accuracy (%)</b>	<b>AUC</b>	<b>P-value</b>
<b>Consensus</b>	87.2%	81.2%	84.6%	0.89	
<b>R1</b>	81.8%	81.0%	81.5%	0.85	0.48
<b>R2</b>	82.9%	77.8%	80.6%	0.87	0.37
<b>R3</b>	87.2%	78.6%	83.6%	0.89	0.62
<b>R4</b>	87.9%	76.0%	82.8%	0.85	0.62



Supplemental Figure 1. Bland-Altman plots for the two selected features between the group consensus contours and individual contours; Blue line is average difference and red dashed lines are 95% limits of agreement.